Journal of Multivariate Analysis **74**, 282–295 (2000)
doi:10.1006/jmva.1999.1887, available online at http://www.idealibrary.com on IDEAL®

Degeneracy in Heteroscedastic Regression Models

Kim-Hung Li and Nai Ng Chan

Department of Statistics, The Chinese University of Hong Kong, Hong Kong
E-mail: khli@cuhk.edu.hk and nnchan@hp735.sta.cuhk.edu.hk

Received February 4, 1999

The maximum likelihood estimation in a regression model with heteroscedastic errors is considered. When the design matrices in the model are inappropriately specified, the maximum likelihood estimates of the variances of certain observations are found to be zero irrespective of the observed values, resulting in degeneracy. Necessary and sufficient conditions for degeneracy are given and used for its

Key words and phrases: degeneracy, generalized linear model with varying dispersion, heteroscedastic error, unbounded likelihood.

1. INTRODUCTION

The study of the estimation in parametric regression models with additive and heteroscedastic errors was launched decades ago. Their induced problems have been constantly of interest in areas such as econometrics [10, Chapter 11] and have been further stimulated by investigations of the Taguchi [19] type in which the importance of variability control in quality improvement is demonstrated. This new development in quality engineering has aroused the interest of statisticians in the in-depth study of the dispersion effects of these models; see [2–5, 11–13, 20].

Let z be an observable response, and let x and y be two design (column) vectors associated with z. A (parametric) heteroscedastic regression model for z is expressible as

$$z = g(x'\beta) + h(y'\gamma)\,\varepsilon,$$

where $g(\cdot)$ and $h(\cdot)$ are known smooth real-valued functions (x' means the transpose of x), ε is a standardized error, and β and γ are unknown parameter vectors. As a generalization, Smyth [18] considered a generalized linear model with varying dispersion. A common choice of $g(\cdot)$ is the identity function and in some research works, the variance $h^2(y'\gamma)$ of

0047-259X/00 $35.00

z is assumed to be proportional to a known function of the mean $g(x'\beta)$. Rutemiller and Bowers [17] considered the case where $g(\cdot)$ and $h(\cdot)$ are both identity functions. Hildreth and Houck [9], in their study of random coefficients linear models, deduced heteroscedastic models with $g(\cdot)$ the identity function and $h(\cdot)$ the square root function. And yet the most popular choice of $h(\cdot)$ is the exponential function [1, 5, 7, 8, 14, 16, 20].

In this paper the method of maximum likelihood estimation is adopted for parameter estimation of the models above due to its great generality and appealing properties. The function $h(\cdot)$ is chosen to be the exponential function, and y may or may not contain a constant term. We take $g(\cdot)$ to be a strictly increasing function from the real line, R, onto R (thus the inverse function $g^{-1}(\cdot)$ exists). Also there may be assumptions on the smoothness of $g(\cdot)$ so as to guarantee that certain asymptotic properties of the maximum likelihood estimator (MLE) hold, though these assumptions are not essential in our subsequent arguments. The error ε is assumed to follow the standard normal distribution.

There are various kinds of non-regularity problems in association with the maximum likelihood estimation. Cheng and Traylor [6] gave a good review on the problems where non-regularity is an inherent characteristic of the model, often causing the likelihood to become unbounded. Non-identifiability is another example; this arises due to improper choice of the values of design variables. Here we consider a degeneracy problem which is caused by designs under which the maximum likelihood estimates of the variances of certain variables become zero no matter what the observed values are. The problem is usually, but not always, accompanied by the unbounded likelihood problem.

From now on, let z_i, $i=1, ..., n$, be independent univariate observations on z. The distribution of z_i depends on two covariate (row) vectors X_i and Y_i of dimensions p and m, respectively, as

$$z_i = g(X_i\beta) + \exp(Y_i\gamma/2)\,\varepsilon_i \qquad \text{for} \quad i=1, ..., n, \tag{1}$$

where ε_i, $i=1, ..., n$, are independent and identically distributed as the standard normal distribution. The variance of z_i given X_i and Y_i, denoted by σ_i^2, is $\exp(Y_i\gamma)$. The model in (1) has two design matrices $X \equiv (X'_1, ..., X'_n)'$ and $Y \equiv (Y'_1, ..., Y'_n)'$ of dimensions $n \times p$ and $n \times m$, respectively. Write $z = (z_1, ..., z_n)'$. In order to have the parameters β and γ identifiable, we assume that both X and Y are of full column rank, and that $n \geqslant p+m$, which is the total number of unknown parameters. If Y is simply an $n \times 1$ matrix of 1's (which will also be denoted by 1), the model in (1) reduces to a regression with homoscedastic error.

In previous works, $g(\cdot)$ in (1) is usually assumed to be the identity function, and Y possesses a column of 1's. Park [16] considered the case in which $m=2$. The model in (1) for general m has been investigated by Harvey [8] and Aitkin [1], who also provided GLIM macros for maximum likelihood estimations. Wang [20] and Chan *et al.* [5] made use of the model to analyze data from Taguchi's experiments.

Write the kernel of the negative log likelihood of (1) as

$$L(\beta, \gamma; z) = 1' Y\gamma + \sum [z_i - g(X_i\beta)]^2 \exp(-Y_i\gamma).$$

Given the observed z, the MLE of the parameter (β, γ) is the parameter value that minimizes $L(\beta, \gamma; z)$. Denote the infimum of $L(\beta, \gamma; z)$ with respect to β and γ by $M(z)$.

EXAMPLE 1. Suppose that six runs are conducted under model (1) with designs

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}$$

and

$$Y' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 0 & -1 & -1 & 0 & 1 \end{bmatrix}.$$

For any $z_1, ..., z_6$, choose β and γ to satisfy $X_5\beta = g^{-1}(z_5)$ and $X_6\beta = g^{-1}(z_6)$, and choose $\gamma = t[-1 \;\; 2 \;\; -1]'$ for a positive constant t. Then

$$L(\beta, \gamma; z) = (z_1 - g(X_1\beta))^2 + (z_2 - g(X_2\beta))^2 \exp(-t) + (z_3 - g(X_3\beta))^2 + (z_4 - g(X_4\beta))^2 - 6t.$$

Clearly $M(z) = -\infty$ by letting t diverge to ∞. In this case, σ_5^2 is approximately estimated by $\exp(Y_5\gamma)$, which is equal to $\exp(-3t)$, and σ_6^2 is approximately estimated by $\exp(Y_6\gamma) = \exp(-4t)$. Both estimates of the variances tend to zero as t diverges to ∞ and are independent of the observations. Thus, standard inference for MLE fails. Also, the estimate is not unique; $L(\beta, \gamma; z)$ also diverges to $-\infty$ when β and γ are such that $X_1\beta = g^{-1}(z_1)$ and $X_2\beta = g^{-1}(z_2)$, and $\gamma = t[-1 \;\; -2 \;\; -1]'$, where t diverges to ∞. The degeneracy can be avoided in the design stage by setting

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix},$$

which disallows, with probability one, the existence of the β estimate above. If the six runs have already been conducted, a remedy for the problem is to perform an additional run at the Y covariate $Y_7 = (1 \;\; 0 \;\; 0)$.

Example 1 demonstrates a serious estimation problem for model (1). It is not easy to discover degeneracy by visual inspection of the design matrices X and Y. This also shows that the problem can be easily removed in the design stage of the experiment when the problem is detected. A formal definition of degeneracy is stated as follows:

DEFINITION 1. Model (1) or its design pair (X, Y) is said to be *degenerate* if for *any* observation z, there exist two sequences $\{\beta_k\}$ and $\{\gamma_k\}$ of vectors such that (i) $\lim_{k\to\infty} L(\beta_k, \gamma_k; z) = M(z)$ $(\equiv \min_{(\beta,\gamma)} L(\beta, \gamma; z))$ and (ii) $\lim_{k\to\infty} Y_i\gamma_k = -\infty$ for at least one i.

DEFINITION 2. A design pair (X, Y) is said to have *unbounded likelihood* if $M(z) = -\infty$ for all z.

It is easy to show that unbounded likelihood pairs are degenerate, but the converse is not necessarily true (see Section 2.2). Since the standard MLE inference fails in a degenerate model, the detection of degeneracy becomes essential in the design stage of an experiment. In what follows, the main results about characterizations of degeneracy are presented in Section 2. We state necessary and sufficient conditions for degeneracy for the case $1'Y \neq 0$. In the less common situation $1'Y = 0$ (which implies that there is no constant term in the Y design matrix), only partial results are available. The proofs of the theorems in Section 2 are given in Section 3. Discussions are made in Section 4.

2. MAIN RESULTS

2.1. *Degeneracy for the Case* $1'Y \neq 0$

THEOREM 1. *For model* (1) *with* $1'Y \neq 0$, *the following three conditions are equivalent*:

(A1) *the model is degenerate*;

(A2) *the design has unbounded likelihood* (*i.e.*, $M(z) = -\infty$ *for all* z);

(A3) *there exists an* $m \times 1$ *vector* b *such that* $1'Yb < 0$ *and* $\{X_i$: *for* i *such that* $Y_ib < 0\}$ *is a set of linearly independent row vectors.*

If $m \geqslant 2$, (A4) *is equivalent to any of* (A1), (A2), *and* (A3), *where*

(A4) *Let* $K = (K_{ij})$ *be any fixed* $n \times m$ *matrix which is obtained by post-multiplying* Y *by a nonsingular square matrix such that the* 1*st component* $(1'K)_1 < 0$ *and the* i*th component* $(1'K)_i = 0$ *for* $i \neq 1$. *There exists a set of indices* $\{i_1, ..., i_{m-1}\} \subseteq \{1, ..., n\}$ *such that the* $(m-1)$ *equations*

$K_{i_k 1} + \sum_{j=2}^{m} K_{i_k j} t_j = 0$, $k = 1, ..., m-1$, *have a unique solution* $(t_2, ..., t_m) = (c_2, ..., c_m)$, *and the set* $\{X_i$: *for* i *such that* $K_{i1} + \sum_{j=2}^{m} K_{ij} c_j < 0\}$ (*which is necessarily nonempty*) *of row vectors is linearly independent.*

Remarks. (1) The case $1'Y \neq 0$ occurs in many applications, in particular where Y contains a column of 1's.

(2) For a degenerate model (1) with $1'Y \neq 0$, the index set $\{i$: $Y_i b < 0\}$ for the b in Condition (A3) is just $\{i$: $\lim_{k \to \infty} Y_i \gamma_k = -\infty\}$ in (ii) of Definition 1 (as seen from the proof of the theorem).

(3) A model (1) with $m = 1$ and $1'Y \neq 0$ is degenerate if either $\{X_i$: for i such that $Y_{i1} < 0\}$ or $\{X_i$: for i such that $Y_{i1} > 0\}$ is a linearly independent set.

(4) If $m = 2$ and $K_{i1} < 0$ for all i, the model is degenerate if, and only if, either $\{X_i$: for i such that $K_{i2} \leqslant 0\}$ or $\{X_i$: for i such that $K_{i2} \geqslant 0\}$ is linearly independent. The necessity part is obvious since when $c_2 > 0(<0)$, $\{X_i$: for i such that $K_{i1} + K_{i2} c_2 < 0\} \supseteq \{X_i$: for i such that $K_{i2} \leqslant 0(\geqslant 0)\}$. The sufficiency follows from choosing i_1 such that $K_{i_1 2} \neq 0$ and $K_{i_1 1}/K_{i_1 2} \leqslant (\geqslant) K_{i1}/K_{i2}$ for all i with $K_{i2} \neq 0$ (i_1 exists as the assumption that Y has full column rank guarantees that the second column is nonzero). We have then $\{X_i$: for i such that $K_{i1} + K_{i2} c_2 < 0\} = \{X_i$: for i such that $K_{i2} \geqslant 0$ $(\leqslant 0)\}$.

(5) The matrix K in (A4) can be obtained from Y by performing a finite number of column operations on Y.

Although Condition (A4) is not as neat as Condition (A3), it is particularly suitable for detecting degeneracy numerically, as demonstrated below.

EXAMPLE 1 (Revisited). For the design Y in Example 1, a corresponding K matrix as considered in (A4) is

$$K' = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 0 & -1 & -1 & 0 & 1 \end{bmatrix}.$$

As the third and the fourth rows of K are identical, we need only to consider all subsets of size 2 from $\{1, 2, 3, 5, 6\}$ in the determination of the (c_2, c_3)'s. Table I lists all cases.

By Theorem 1, the model is degenerate if, and only if, at least one of the following sets is independent:

$$\{X_1, X_2\}, \quad \{X_2, X_3, X_4\}, \quad \{X_3, X_4, X_5\}, \quad \{X_5, X_6\}. \tag{2}$$

TABLE I

(c_2, c_3) and the Corresponding Rows of X

Subset of size 2	(c_2, c_3)	$\{X_i : K_{i1} + \sum_{j=2}^{6} K_{ij} c_j < 0\}$
1, 2	(0, 1)	$\{X_3, X_4, X_5, X_6\}$
1, 3	(2, −1)	$\{X_5, X_6\}$
1, 5	(−1, 2)	$\{X_2, X_3, X_4\}$
1, 6	(0, 1)	$\{X_2, X_3, X_4, X_5\}$
2, 3	(1, −1)	$\{X_1, X_5, X_6\}$
2, 5	No solution	
2, 6	(1, 2)	$\{X_3, X_4, X_5\}$
3, 5	(−1, −1)	$\{X_1, X_2, X_6\}$
3, 6	(−2, −1)	$\{X_1, X_2\}$
5, 6	(−1, 0)	$\{X_1, X_2, X_3, X_4\}$

Note that we need not check the independence of all sets (e.g., $\{X_3, X_4, X_5, X_6\}$) in the third column of Table I as the independence of each ignored set implies the independence of at least one of the sets in (2). After inspecting the sets in (2), it is easy to see why the model (with the given Y) is degenerate if

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix},$$

and why it is not if

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}.$$

Also recall that an additional run with $Y_7 = (1\ 0\ 0)$ helps in removing degeneracy. This is because the new K is

$$K' = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 0 & 0 & -1 & -1 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 & 0 \end{bmatrix},$$

and that $K_{71} + K_{72} c_2 + K_{73} c_3 < 0$ for all c_2 and c_3 implies that X_7 has to be added to all four sets in (2).

If Condition (A4) is used to detect degeneracy, we have to solve ${}_nC_{m-1}$ systems of $(m-1)$ equations. Each system produces at most one $(c_2, ..., c_m)$ (we ignore systems that give no solution or more than one solution). Suppose p and m are held fixed, while n tends to ∞. Solving a system of $(m-1)$ equations needs $O(1)$ time. Checking for signs of $K_{i1} + \sum_{j=2}^{m} K_{ij} c_j$

for $i=1, \ldots, n$ requires $O(n)$ time. Checking whether the selected rows of X form an independent set of vectors requires $O(1)$ time because if the number of selected rows is larger than p, the answer must be negative. Therefore, the overall time for detecting degeneracy using Condition (A4) is $O(n^m)$ which is within a polynomial time.

All subsets of size $(m-1)$ of $\{1, \ldots, n\}$ can be ordered in a sequence so that successive subsets differ only by one element [15, pp. 26–38]. Solving systems of equations selected in this order should be beneficial as the computation of the solution in a system of equations would help solving the subsequent system. Specifically, let U and V be two $(m-1)\times(m-1)$ nonsingular matrices which differ only by their last rows. Write

$$V=\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix},$$

where V_{11} is $(m-2)\times(m-2)$, V_{12} is $(m-2)\times 1$, V_{21} is $1\times(m-2)$, and V_{22} is 1×1. Denote

$$U^{-1}=\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \quad \text{and} \quad V^{-1}=\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix},$$

where both C_{22} and D_{22} are 1×1. If $(V_{21}C_{12}+V_{22}C_{22})\neq 0$ and $C_{22}\neq 0$, then given V and U^{-1}, V^{-1} can be computed as

$$D_{22}=\rho C_{22},$$

$$D_{12}=\rho C_{12},$$

$$D_{11}=L-\rho C_{12}V_{21}L,$$

and

$$D_{21}=-\rho C_{22}V_{21}L,$$

where $\rho=1/(V_{21}C_{12}+V_{22}C_{22})$, and $L=C_{11}-C_{12}C_{21}/C_{22}$.

2.2. *Degeneracy for the Case* $1'Y=0$

THEOREM 2. *Model* (1) *with* $1'Y=0$ *is degenerate with* $M(z)=0$ *for all* z *if, and only if, there exists a nonzero vector* b *such that the set* $\{X_i$: *for* i *such that* $Y_i b\leqslant 0\}$ *is linearly independent.*

THEOREM 3. *Suppose that model* (1) *with* $1'Y=0$ *is degenerate. Then there exists a nonzero vector* b *such that the set* $\{X_i$: *for* i *such that* $Y_i b<0\}$ *is linearly independent.*

The necessary condition in Theorem 3 is not a sufficient condition for degeneracy as shown below.

EXAMPLE 2. Suppose that $n=2$, and $z_1 \sim N(\mu, \sigma^2)$ and $z_2 \sim N(\mu, 1/\sigma^2)$ are independent normal variates. When expressed in the form of model (1), the function $g(\cdot)$ is the identity function, $X=(1, 1)'$ and $Y=(1, -1)'$. It is easy to see from Theorem 2 that the model is degenerate. By choosing $\mu=z_1$ and $\sigma=0$, $L(\beta, \gamma; z)$ attains its global minimum 0. Suppose one more run is conducted with $z_3 \sim N(\mu, 1)$. Then $X=(1, 1, 1)'$ and $Y=(1, -1, 0)'$. Theorem 2 shows that the model cannot be degenerate with $M(z)=0$. However, by Theorem 3, it satisfies a necessary condition for degeneracy. If it is degenerate, $M(z)$ must be either $(z_3-z_1)^2$ or $(z_3-z_2)^2$. For $z=(1, 2, 12)'$, the minimum of $L(\beta, 0; z)$ is less than $(z_3-z_1)^2$ and $(z_3-z_2)^2$. Therefore, at least for this z, the MLE of σ is not 0 or ∞, contradicting our assumption. Thus the model with one more run is not degenerate, and so the necessary condition in Theorem 3 is not sufficient for degeneracy.

3. PROOFS OF THE THEOREMS

3.1. *Lemmas*

LEMMA 1. *Let $b_1, b_2, ..., b_r$ be s-dimensional row vectors, and $q_1, q_2, ...$ be a sequence of s-dimensional column vectors such that for each i $(1 \leqslant i \leqslant r)$, $\lim_{k \to \infty} b_i q_k = \zeta_i$, where $-\infty \leqslant \zeta_i \leqslant \infty$. Then there exist s-dimensional column vectors c and v such that*

(i) *$b_i c = \zeta_i$ and $b_i v = 0$ if ζ_i is finite;*

(ii) *$b_i c > 0$ and $b_i v > 0$ if $\zeta_i = \infty$; and*

(iii) *$b_i c < 0$ and $b_i v < 0$ if $\zeta_i = -\infty$.*

Proof. Part 1: Let η be a positive value larger than $|\zeta_i|$ for all finite ζ_i. Denote by R^s the space of s-dimensional column vectors and define $\varphi(u)$ to be the function from R^s to R^r such that for $i=1, ..., r$

$$\begin{aligned} \varphi(u)_i &= b_i u && \text{if} \quad |b_i u| < \eta \\ &= \eta && \text{if} \quad b_i u \geqslant \eta \\ &= -\eta && \text{if} \quad b_i u \leqslant -\eta. \end{aligned}$$

As $\varphi(R^s)$ is a closed subset in R^r, and $\varphi(q_k)$ has a finite limit, there exists a vector c in R^s such that $\varphi(c) = \lim_{k \to \infty} \varphi(q_k)$. If ζ_i is finite, then $\varphi(c) = \lim_{k \to \infty} \varphi(q_k) = \zeta_i \neq \pm\eta$ and thus $b_i c = \zeta_i$. If $\zeta_i = \infty$, then $\varphi(c) = \lim_{k \to \infty} \varphi(q_k) = \eta$ implying that $b_i c \geqslant \eta > 0$. Similarly, we have $b_i c \leqslant -\eta < 0$, if $\zeta_i = -\infty$. These prove the existence of c in the lemma.

Part 2: Let $\{w_k\}$ be a subsequence of $\{q_k\}$ such that if $\lim_{k\to\infty} b_i q_k = \pm\infty$, the successive differences $b_i w_k - b_i w_{k-1}$ diverge to $\pm\infty$ as k tends to infinity. If ζ_i is finite, $b_i(w_k - w_{k-1})$ has zero limit. Applying Part 1 to the sequence $\{w_k - w_{k-1}\}$, we prove the existence of v. ∎

LEMMA 2. *Suppose that for every z, there exist an $m\times 1$ vector c and a $p\times 1$ vector λ such that*

(i) $\{i: Y_i c < (\leqslant)\, 0\} \neq \phi$,

(ii) $g^{-1}(z_i) = X_i\lambda$ *for all* $i \in \{i: Y_i c < (\leqslant)\, 0\}$, *and*

(iii) *if* $z_i = g(0)$ *for all* i *and* $1'Y \neq 0$, *then* $1'Yc < 0$.

Then there is a nonzero vector b such that the set $\{X_i$: for i such that $Y_i b < (\leqslant)\, 0\}$ is nonempty and linearly independent. Moreover if $1'Y \neq 0$, we have $1'Yb < 0$.

Proof. Write $g^{-1}(z) = (g^{-1}(z_1), ..., g^{-1}(z_n))'$. For each non-empty subset S of $\{1, ..., n\}$, define $\Gamma(S) = \{g^{-1}(z)$: there is a vector c defined in the Lemma such that $\{i: Y_i c < (\leqslant)\, 0\} = S\}$. As every $g^{-1}(z)$ must belong to at least one $\Gamma(S)$, we have $\bigcup \Gamma(S) = R^n$. From (ii), $\Gamma(S) \subseteq \{g^{-1}(z)$: there exists a λ such that $X_i\lambda = g^{-1}(z_i)$ for all $i \in S\}$. Therefore, $\Gamma(S)$ is a subset of a subspace of dimension at most n. As the number of subsets S is finite and $\bigcup \Gamma(S) = R^n$, there is at least one S, say S^*, such that $\Gamma(S^*) = R^n$. This implies that $\{g^{-1}(z)$: there exists a λ such that $X_i\lambda = g^{-1}(z_i)$ for all $i \in S^*\} = R^n$, and thus $\{X_i : i \in S^*\}$ is a set of linearly independent vectors. For the special case where $z_i = g(0)$ for all i, let b be the corresponding vector c as defined in the Lemma such that $\{i: Y_i c < (\leqslant)\, 0\} = S^*$ (noting that $0 \in R^n = \Gamma(S^*)$). Then $\{X_i$: for i such that $Y_i b < (\leqslant)\, 0\}$ is nonempty and independent. b cannot be a zero vector as $\{i: Y_i b < (\leqslant)\, 0\}$ must be a proper subset of $\{1, ..., n\}$ (since we assume that $p < n$). If $1'Y \neq 0$, from (iii) we have $1'Yb < 0$, completing the proof. ∎

3.2. *Proof of Theorem* 1

For model (1) with $1'Y \neq 0$, it is obvious that Condition (A2) implies (A1). To show that (A3) implies (A2), choose $\gamma_k = kb$, and select β to satisfy $X_i\beta = g^{-1}(z_i)$ for all i with $Y_i b < 0$ on noting that the existence of β is guaranteed by Condition (A3). As $\lim_{k\to\infty} L(\beta, \gamma_k; z) = -\infty$, we have $M(z) = -\infty$ for all z. This also shows the validity of Remark (2) in Section 2.1. The proof for (A1) implying (A3) requires the following.

LEMMA 3. *For a degenerate model* (1) *with* $1'Y \neq 0$, *for each z there exist an $m\times 1$ vector c and a $p\times 1$ vector λ such that* (i) $\{i: Y_i c < 0\} \neq \phi$, *and* (ii) $g^{-1}(z_i) = X_i\lambda$ *for all i such that* $Y_i C < 0$.

Proof. Since the model is degenerate, for any z there exist $\{\beta_k\}$ and $\{\gamma_k\}$ such that $\lim_{k\to\infty} L(\beta_k, \gamma_k; z) = M(z)$, and $S = \{i : \lim_{k\to\infty} Y_i\gamma_k = -\infty\} \neq \phi$. By choosing suitable subsequences of $\{\beta_k\}$ and $\{\gamma_k\}$, we may, without loss of generality, assume that all of the following limits exist (which might be $\pm\infty$):

$$\lim_{k\to\infty} X_i\beta_k, \qquad \lim_{k\to\infty} Y_i\gamma_k, \qquad \lim_{k\to\infty} 1'Y\gamma_k,$$

$$\lim_{k\to\infty} \exp(-Y_i\gamma_k)/(-1'Y\gamma_k), \qquad \text{and} \qquad \lim_{k\to\infty} Y_i\gamma_k/\log(|1'Y\gamma_k|).$$

By Lemma 1, there exists a vector λ such that $X_i\lambda = \lim_{k\to\infty} X_i\beta_k$ if $\lim_{k\to\infty} X_i\beta_k$ is finite, $X_i\lambda > 0$ if $\lim_{k\to\infty} X_i\beta_k = \infty$, and $X_i\lambda < 0$ if $\lim_{k\to\infty} X_i\beta_k = -\infty$. Write $S^* = \{i : \lim_{k\to\infty} Y_i\gamma_k = -\infty$, and $\lim_{k\to\infty} X_i\beta_k \neq g^{-1}(z_i)\}$.

Case 1: $S^* = \phi$. From Lemma 1 with $r = n$ and $b_i = Y_i$ for all i, there exists a vector c such that $\{i : Y_i c < 0\} = S$. As $S^* = \phi$, for any $i \in S$, $\lim_{k\to\infty} X_i\beta_k = g^{-1}(z_i)$, and thus $X_i\lambda = g^{-1}(z_i)$.

Case 2: $S^* \neq \phi$. As $M(z)$ cannot be ∞, we have

(i) $\lim_{k\to\infty} 1'Y\gamma_k = -\infty$ (since $S^* \neq \phi$); and

(ii) if $\lim_{k\to\infty} \exp(-Y_i\gamma_k)/(-1'Y\gamma_k) = \infty$ (i.e., $\lim_{k\to\infty}[Y_i\gamma_k + \log(-1'Y\gamma_k)] = -\infty$), then $\lim_{k\to\infty} X_i\beta_k = g^{-1}(z_i)$.

Define a sequence of $(m+1)$-dimensional vectors $q_k = \gamma_k/\log(-1'Y\gamma_k)$ for $k = 1, 2, \ldots,$ and let $b_i = Y_i$ for $i = 1, \ldots, n$, and $b_{n+1} = 1'Y$. From Lemma 1, there exists v such that $b_i v < 0$ if and only if $\lim_{k\to\infty} b_i q_k = -\infty$, and $b_i v = 0$ if and only if $\lim_{k\to\infty} b_i q_k$ is finite. Since $\lim_{k\to\infty} b_{n+1} q_k = -\infty$, $1'Yv < 0$. As a consequence, $\{i : Y_i v < 0\} \neq \phi$. If $Y_i v < 0$, then $\lim_{k\to\infty} b_i q_k = -\infty$. It follows that $\lim_{k\to\infty} Y_i\gamma_k/\log(-1'Y\gamma_k) = -\infty$. Therefore, $\lim_{k\to\infty}[Y_i\gamma_k + \log(-1'Y\gamma_k)] = -\infty$. Using (ii), we have then $\lim_{k\to\infty} X_i\beta_k = g^{-1}(z_i)$, and so we have $g^{-1}(z_i) = X_i\lambda$. By setting $c = v$, the result follows. ∎

If $1'Y \neq 0$ and $z_i = g(0)$ for all i, take $\lambda = 0$ and c be such that $1'Yc < 0$. Clearly the conditions (using the "$<$" sign) in Lemma 2 for this special c holds. When model (1) is degenerate, Lemma 3 guarantees the conditions (using the "$<$" sign) in Lemma 2 for all other z. Applying Lemma 2, we have (A1) implies (A3).

In order to prove the equivalence of (A3) and (A4) for the case $m \geqslant 2$ and $1'Y \neq 0$, let us write $S[b] = \{i : Y_i b < 0\}$ for an $m \times 1$ vector b, and $B = \{b : 1'Yb < 0\}$. Thus Condition (A3) can be restated as follows:

(A3) There exists $b \in B$ such that $\{X_i : i \in S[b]\}$ is independent.

To prove the equivalence, we need the following.

LEMMA 4. *For any $b \in B$ such that the set $\{1'Y\} \cup \{Y_i\colon Y_i b = 0\}$ spans a subspace of dimension $r < m$, there is a $b^* \in B$ such that $S[b^*] \subseteq S[b]$, and $\{1'Y\} \cup \{Y_i\colon Y_i b^* = 0\}$ spans a subspace of dimension at least $(r+1)$.*

Proof. Since Y is of rank m, there is a vector Y_j which is not a linear combination of $\{1'Y\} \cup \{Y_i\colon Y_i b = 0\}$. So for any b and i with $Y_i b = 0$, there is a vector b^* such that $1'Yb^* = 1'Yb$, $Y_j b^* = 0$ and $Y_i b^* = 0$. In case that $S[b^*] \subseteq S[b]$ does not hold, there is an integer k such that $Y_k b^* < 0$ and $Y_k b > 0$. Write

$$\varepsilon = \min_i \{ -Y_i b / Y_i b^* : (Y_i b)(Y_i b^*) < 0\}$$

and let the minimum be attained at $i = u$. Clearly, $(b + \varepsilon b^*) \in B$, and

(i) $Y_i b < 0$ implies that $Y_i(b + \varepsilon b^*) \leqslant 0$;

(ii) $Y_i b = 0$ implies that $Y_i(b + \varepsilon b^*) = 0$; and

(iii) $Y_i b > 0$ implies that $Y_i(b + \varepsilon b^*) \geqslant 0$.

Therefore, $S[b + \varepsilon b^*] \subseteq S[b]$, and obviously $Y_u(b + \varepsilon b^*) = 0$ and $Y_u b \neq 0$. To prove that $\{1'Y\} \cup \{Y_i\colon Y_i(b + \varepsilon b^*) = 0\}$ spans a subspace of dimension at least $r+1$, it suffices to show that Y_u cannot be spanned by $\{1'Y\} \cup \{Y_i\colon Y_i b = 0\}$. For otherwise, $Y_u = \alpha 1'Y + \sum_{i\colon Y_i b = 0} \eta_i Y_i$. However, since $1'Y(b + \varepsilon b^*) < 0$ and $Y_i(b + \varepsilon b^*) = 0$ whenever $Y_i b = 0$, we then have $\alpha = 0$, which contradicts the fact that $Y_u b \neq 0$. This completes the proof. ∎

From Lemma 4, (A3) is equivalent to

(A3*) There is a $b \in B$ such that $\{1'Y\} \cup \{Y_i\colon Y_i b = 0\}$ spans a subspace of dimension m and $\{X_i : i \in S[b]\}$ is linearly independent.

(A3) is again equivalent to

(A3**) There is a unique (up to a positive factor) vector b in B such that $Y_i b = 0$ for exactly $(m-1)$ i's, and $\{X_i : i \in S[b]\}$ is independent.

Also, Condition (A4) follows from (A3**) by observing that (i) $K = YD$ for a nonsingular matrix D with $(1'K)_1 < 0$ and $(1'K)_i = 0$ for all $i \neq 1$, (ii) $1'Yb < 0$ is equivalent to $1'K(D^{-1}b) < 0$, that is, $(D^{-1}b)_1 > 0$, and (iii) we can rescale b so that $(D^{-1}b)_1 = 1$ and thus b becomes unique as required in (A4). That (A4) implies (A3**) follows by setting $b = D(1, c_2, ..., c_m)'$.

3.3. *Proof of Theorem* 2

Sufficiency: As the vector $b \neq 0$, $\{i: Y_i b < 0\} \neq \phi$. Write $\gamma_k = kb$ for each positive integer k, and choose β to satisfy $X_i \beta = g^{-1}(z_i)$ whenever $Y_i b \leqslant 0$. Then $\lim_{k\to\infty} L(\beta, \gamma_k; z) = 0$, which is $M(z)$. Moreover, $\lim_{k\to\infty} Y_i \gamma_k = -\infty$ for i such that $Y_i b < 0$.

Necessity: For any given z, let $U = \{i: \lim_{k\to\infty} Y_i \gamma_k \leqslant \infty\}$, which is necessarily nonempty. Since $M(z) = 0$, we have for all $i \in U$, $\lim_{k\to\infty} X_i \beta_k = g^{-1}(z_i)$. From Lemma 1, there exist c and λ such that $U = \{i: Y_i c \leqslant 0\}$ and $X_i \lambda = g^{-1}(z_i)$ for all $i \in U$. The proof is complete by applying Lemma 2 (using the "$\leqslant$" sign). ▮

3.4. *Proof of Theorem*

For any z with the corresponding γ_k in Definition 1, write $S = \{i: \lim_{k\to\infty} Y_i \gamma_k = -\infty\}$, which is necessarily nonempty by the assumption. For all $i \in S$, we have $\lim_{k\to\infty} X_i \beta_k = g^{-1}(z_i)$, for otherwise, $M(z) = \infty$, rendering an impossible situation. By Lemma 1, there exist c and λ such that $S = \{i: Y_i c < 0\}$ and for all $i \in S$, $X_i \lambda = g^{-1}(z_i)$. The theorem then follows from Lemma 2 (using the "$<$" sign).

4. DISCUSSION

Degeneracy in a heteroscedastic regression model (as in (1)) is a serious problem in parameter estimation. Equivalent conditions for degeneracy are given in Section 2 for the case $1'Y \neq 0$. These conditions should be checked prior to the data collection stage, especially when the number of experimental runs is small and the elements of the design matrices take discrete values. Condition (A4) provides a simple and efficient algorithm for numerical checking of degeneracy in polynomial time. Remarks (3) and (4) in Section 2 simplify the condition in special situations.

If there is no tight restriction on the number of experimental runs, a simple way to avoid degeneracy is to have replicates at each of the design points. For example, if the second run is a replicate of the first run, the estimate of σ_1 would not be zero provided that $z_1 \neq z_2$.

Empirically the following phenomena in the iterative process of minimizing the function $L(\beta, \gamma; z)$ (of Section 1) indicate degeneracy.

(i) Certain residuals, $z_i - g(X_i \beta_k)$, are close to zero;

(ii) $L(\beta_k, \gamma_k; z)$ diverges to $-\infty$ (if $1'Y \neq 0$); and

(iii) $\hat{\sigma}_i$ tends to zero for some i.

These phenomena usually occur simultaneously for the case $1'Y \neq 0$, and their occurrence indicates a necessity for theoretical checking of degeneracy by using the Theorems in Section 2. A preliminary attempt could be to check Condition (A3) for b taking the value γ_k (normalized to have norm 1) for a sufficiently large k.

ACKNOWLEDGMENTS

The authors thank the referee for his helpful comments and suggestions. This research was supported by a direct grant (Project ID: 2060114) from the Chinese University of Hong Kong.

REFERENCES

1. M. Aitkin, Modelling variance heterogeneity in normal regression using GLIM, *Appl. Statist.* **36** (1987), 332–339.
2. G. E. P. Box, Signal-to-noise ratios, performance criteria and transformations (with discussion), *Technometrics* **30** (1988), 1–40.
3. G. E. P. Box, Quality improvement—The new industrial revolution, *Internat. Statist. Rev.* **61** (1993), 3–19.
4. G. E. P. Box and R. D. Meyer, Dispersion effects from fractional designs, *Technometrics* **28** (1986), 19–27.
5. N. N. Chan, K. H. Li, and K. Sharpe, Analysing Taguchi's experiments using GLIM, *Total Quality Management* **3** (1992), 189–196.
6. R. C. H. Cheng and L. Traylor, Non-regular maximum likelihood problems, *J. Roy. Statist. Soc. Ser. B* **57** (1995), 3–44.
7. R. D. Cook and S. Weisberg, Diagnostics for heterscedasticity in regression, *Biometrika* **70** (1983), 1–10.
8. A. C. Harvey, Estimating regression models with multiplicative heteroscedasticity, *Econometrica* **44** (1976), 461–465.
9. C. Hildreth and J. P. Houck, Some estimators for a linear model with random coefficients, *J. Amer. Statist. Assoc.* **63** (1968), 584–595.
10. G. G. Judge, W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee, "The Theory and Practice of Econometrics," 2nd ed., Wiley, New York, 1985.
11. R. Léon, A. C. Shoemaker, and R. Kacker, Performance measures independent of adjustment (with discussion), *Technometrics* **29** (1987), 253–286.
12. V. N. Nair, Testing in industrial experiments with ordered categorical data (with discussion), *Technometrics* **28** (1986), 283–311.
13. V. N. Nair and D. Pregibon, A data analysis strategy for quality engineering experiments, *AT&T Tech. J.* **65** (1986), 73–84.
14. V. N. Nair and D. Pregibon, Analyzing dispersion effects from replicated factorial experiments, *Technometrics* **30** (1988), 247–257.
15. A. Nijenhuis and H. S. Wilf, "Combinatorial Algorithms for Computers and Calculators," 2nd ed., Academic Press, New York, 1978.
16. R. E. Park, Estimation with heteroscedastic error terms, *Econometrica* **34** (1966), 888.
17. H. C. Rutemiller and D. A. Bowers, Estimation in a heteroscedastic regression model, *J. Amer. Statist. Assoc.* **63** (1968), 552–557.

18. G. K. Smyth, Generalized linear models with varying dispersion, *J. Roy. Statist. Soc. Ser. B* **51** (1989), 47–60.
19. G. Taguchi, "Introduction to Quality Engineering," Asian Productivity Organization, Tokyo, 1986.
20. P. C. Wang, Tests for dispersion effects from orthogonal arrays, *Comput. Statist. Data Anal.* **8** (1989), 109–117.